**Statistical Inference Final Project**

**PART1: SIMULATION EXERCISE**

TASK:

Investigate exponential distribution in R and compare it with Central Limit theorem .The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

> set.seed(31) # set seed for reproducability

> n <- 40 # 40 samples

> simulations <- 1000 # 1000 simulations

> lambda <- 0.2 # set lambda to 0.2

> simulated\_exponentials <- replicate(simulations, rexp(n, lambda)) # simulate

> means\_exponentials <- apply(simulated\_exponentials, 2, mean) # calculate mean of exponentials

Q1 : Show the sample mean and compare it to the theoretical mean of the distribution.

Sample Mean: 4.9938

Theoretical Mean: 5

> # distrribution mean

> analytical\_mean <- mean(means\_exponentials)

> analytical\_mean

[1] 4.993867

> # analytical mean

> theory\_mean <- 1/lambda

> theory\_mean

[1] 5

Visualization

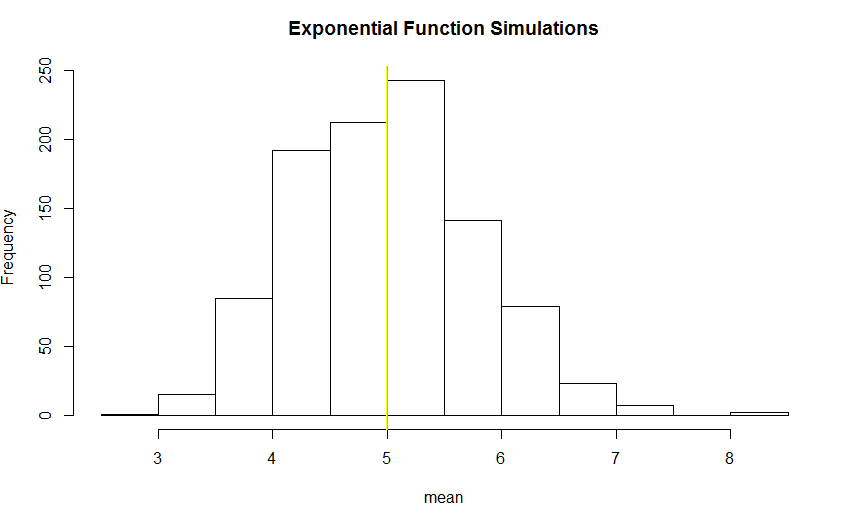
The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution as seen below

> # visualization

> hist(means\_exponentials, xlab = "mean", main = "Exponential Function Simulations")

> abline(v = analytical\_mean, col = "yellow")

> abline(v = theory\_mean, col = "orange")



Q2: Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

Standard Deviation of the distribution is 0.7931608 with the theoretical SD calculated as 0.7905694. The Theoretical variance is calculated as ((1/lambda) \*(1/sqrt(n))) ^2 = 0.625. The actual variance of the distribution is 0. 6291041.Below is the code.

> # standard deviation of distribution

> standard\_deviation\_dist <- sd(means\_exponentials)

> standard\_deviation\_dist

[1] 0.7931608

> # standard deviation from analytical expression

> standard\_deviation\_theory <- (1/lambda)/sqrt(n)

> standard\_deviation\_theory

[1] 0.7905694

> # variance of distribution

> variance\_dist <- standard\_deviation\_dist^2

> variance\_dist

[1] 0.6291041

> # variance from analytical expression

> variance\_theory <- ((1/lambda)\*(1/sqrt(n)))^2

> variance\_theory

[1] 0.625

Q3: Show that the distribution is approximately normal.

Due to central limit theorem the distribution of 40 exponentials should be approximately normal.

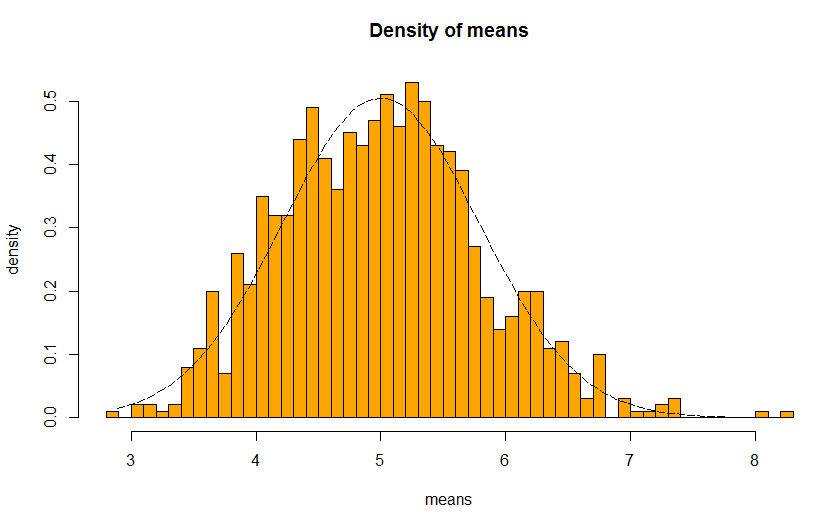
The following visualization portrays this

> xfit <- seq(min(means\_exponentials), max(means\_exponentials), length=100)

> yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(n)))

> hist(means\_exponentials,breaks=n,prob=T,col="orange",xlab = "means",main="Density of means",ylab="density")

> lines(xfit, yfit, pch=22, col="black", lty=5)



The sample quantile plot is seen below

> qqnorm(means\_exponentials)

> qqline(means\_exponentials, col = 2)

\